Stochastic Processes and their Applications 5 (1977) 207–211.

ON THE POSTERIOR DISTRIBUTION OF A DIRICHLET PROCESS GIVEN RANDOMLY RIGHT CENSORED OBSERVATIONS*

J. BLUM and V. SUSARLA

National Science Foundation, and Department of Mathematics, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, U.S.A.

Received 27 October 1976

The purpose of this note is to show that the conditional distribution of a Dirichlet process P given n independent observations $X_1, \dots, X_n$ from P and belonging to measurable sets $A_1, \dots, A_n$ with $A_i \subseteq A_{i+1}$ for $i = 1, \dots, n-1$ is a mixture of Dirichlet processes as introduced by Antoniak. It is also shown that this result is applicable in Bayesian decision problems concerning a random survival distribution under Dirichlet process priors.

Dirichlet process	posterior distributions
mixtures	randomly censored data
random measures	survival distributions
nonparametric	

1. Introduction and summary

Following the notation of Antoniak [1] and Ferguson [2], let the random probability measure P be a Dirichlet process on $(\Theta, \mathscr{A})$ with parameter α, a finite positive measure on $\mathscr{A}$. Let $\theta_1, \dots, \theta_k$ be a random sample from P, $A_i \in \mathscr{A}$ for $i = 1, \dots, k$ with $A_i \subseteq A_{i+1}$ for $i = 1, \dots, k-1$. The purpose of this note is to show (see Theorem 1) that the conditional distribution (c.d.) of P given $X_i \in A_i$ for $i = 1, \dots, k$ is a mixture of Dirichlet processes thus generalizing Theorem 1 of Antoniak [1] to random samples of any size in some situations. Such an extension is shown to be very useful in the Bayesian analysis of randomly censored data from a Dirichlet process in the final section. For any $A_1, \dots, A_k$ in $\mathscr{A}$, we state a result (see Theorem 2) similar to our Theorem 1.

Throughout, $I_A(u) = 1$ if $u \in A$ and $= 0$ if $u \notin A$ where $(u, A) \in \Theta \times A$. For various definitions, we refer the reader to Antoniak [1] or Ferguson [2].

* Research sponsored in part by the United States Army under Contract No. DAAG29-75-C-0024 at the Mathematics Research Center, Madison, by the NSF under Grant No. MCS76-05952, and by a grant from NIH of DHEW under Grant No. 5 R01 GM 23129.

2. Main results

In this section, we prove the following extension to Theorem 1 of Antoniak [1].

Theorem 1. *The conditional distribution of P given $x_i \in A_i$ $(\in \mathscr{A})$ for $i = 1, \dots, k$ with $A_1 \subseteq A_2 \cdots \subseteq A_k$ and $\alpha(A_1) > 0$ is a mixture of Dirichlet processes with transition measure defined by $\alpha_k(u, A) = \alpha(A) + \sum_{i=1}^{k-1} \mu_i(A) + I_A(u)$ for $(u, A) \in \Theta \times \mathscr{A}$ and with mixing measure μ_k where $\mu_1, \dots, \mu_k$ are defined by $\alpha(A_1)\mu_1(A) = \alpha(A \cap A_1)$ and for A in $\mathscr{A}$,*

$$\mu_l(A) = \frac{\alpha(A \cap A_l \cap \bar{A}_{l-1})}{\alpha(A_l) + l - 1} + \sum_{j=1}^{l-1} \frac{\alpha(A \cap A_j \cap \bar{A}_{j-1})}{\alpha(A_j) + j - 1} \prod_{i=j}^{l-1} \left\{ \frac{\alpha(A_i) + i}{\alpha(A_{i+1}) + i} \right\} \tag{1}$$

for $l = 2, \dots, k$ where $\bar{B}$ stands for the complement of B and $A_0 = \emptyset$.

Proof. By Theorem 1 of Antoniak [1], the c.d. of P given $X_1 \in A_1$ is a mixture of Dirichlet processes with transition measure α_1 and mixing measure μ_1 given by

$$\alpha_1(u_1, A) = \alpha(A) + I_A(u_1) \qquad \text{for } (u_1, A) \in \Theta \times \mathscr{A}, \tag{2}$$

and

$$\alpha(A_1)\mu_1(A) = \alpha(A_1 \cap A) \qquad \text{for } A \in \mathscr{A}. \tag{3}$$

To obtain the c.d. of P given $X_1 \in A_1$ and $X_2 \in A_2$, we assume that we have a sample X_2 of size 1 from the mixture of Dirichlet processes described via (2) and (3) and that $X_2 \in A_2$. So, accordingly, let u_1 be a point in A_1 chosen randomly according to μ_1. Now an application of Theorem 1 of Antoniak [1] with $\alpha(\cdot) + I_.(u_1)$ as the parameter of the Dirichlet process and with the observation belonging to A_2 gives the conditional (given u_1) transition and mixing measures defined by

$$\alpha_2(u_2, A \mid u_1) = \alpha(A) + I_A(u_1) + I_A(u_2), \tag{4}$$

and

$$\mu_2(A \mid u_1) = \frac{\alpha(A \cap A_2) + I_{A \cap A_2}(u_1)}{\alpha(A_2) + I_{A_2}(u_1)} \tag{5}$$

for $(u, A) \in \Theta \times \mathscr{A}$. Since u_1 is distributed according to μ_1 with support $\subseteq A_1 \subseteq A_2$, (4) and (5) show that the c.d. of P given $X_1 \in A_1$ and $X_2 \in A_2$ is a mixture of Dirichlet processes with transition and mixing measures defined by

$$\alpha_2(u_2, A) = \alpha(A) + \mu_1(A) + I_A(u_2), \tag{6}$$

and

$$\mu_2(A) = \frac{\alpha(A \cap A_2) + \mu_1(A \cap A_2)}{\alpha(A_2) + 1} = \frac{\alpha(A \cap A_2) + \mu_1(A)}{\alpha(A_2) + 1} \tag{7}$$

for $(u, A) \in \Theta \times \mathscr{A}$.

Proceeding as described above, one obtains that the c.d. of P given $X_1 \in A_1, X_2 \in A_2, \dots, X_l \in A_l$ is a mixture of Dirichlet processes with

$$\alpha_l(u_l, A) = \alpha(A) + \mu_1(A) + \cdots + \mu_{l-1}(A) + I_A(u_l), \tag{8}$$

and

$$\mu_l(A) = \frac{\alpha(A \cap A_l) + \mu_1(A) + \cdots + \mu_{l-1}(A)}{\alpha(A_l) + l - 1} \tag{9}$$

for $(u_l, A) \in \Theta \times \mathscr{A}$ and $l = 2, \ldots, k$. Eq. (8) with $l = k$ gives the transition measure stated in the result. The desired expression for μ_l (see (1)) can be obtained by first solving the linear difference equations $b_l a_l = c_l + \sum_{i=0}^{l-1} a_i$ for $l = 1, 2, \ldots$ for a_l in terms of $a_1, b_1, \ldots, b_l$ and $c_1, \ldots, c_l$ whenever $a_0 = 0$, $c_1 = a_1$, $b_1 = 1$ and $b_i > 0$ for all i. Such a solution leads to

$$a_l = \frac{c_l - c_{l-1}}{b_l} + \sum_{j=1}^{l-1} \frac{c_j - c_{j-1}}{b_j} \prod_{i=j}^{l-1} \left\{ \frac{1 + b_i}{b_{i+1}} \right\} \tag{10}$$

for $l = 2, \ldots$. Substituting $a_i = \mu_i(A)$, $b_i = \alpha(A_i) + i - 1$, and $c_i = \alpha(A \cap A_i)$, we obtain (1) from (10) completing the proof of the result.

Theorem 1 stated above can be generalized as described in the following theorem whose proof follows from the straightforward definition of a conditional distribution.

Theorem 2. *The conditional distribution of P given $X_i \in A_i$ $(\in \mathscr{A})$ with $\alpha(A_i) > 0$ for $i = 1, \ldots, k$ is a mixture of Dirichlet processes with transition measure defined by $\alpha_k((u_1, \ldots, u_k), A) = \alpha(A) + \sum_{i=1}^{k} I_A(u_i)$ for $(u_1, \ldots, u_k) \in \mathsf{X}_1^k \Theta$ and $A \in \mathscr{A}$ and with mixing measure μ_k defined by $\alpha_{(X_1, \ldots, X_k)}(A_1 \times \cdots \times A_k)\mu_k(B) = \alpha_{(X_1, \ldots, X_k)}(B \cap (A_1 \times \cdots \times A_k))$ for B in $\sigma(\mathsf{X}_1^k \mathscr{A})$, the σ-field generated by the field $\mathsf{X}_1^k \mathscr{A}$ and $\alpha_{(X_1, \ldots, X_k)}$ is the marginal probability measure of $(X_1, \ldots, X_k)$.*

Remark. The difference between Theorems 1 and 2 is that the mixing distribution in Theorem 1 is just defined on $\mathscr{A}$ while in the other case, it is on $\sigma(\mathsf{X}_1^k \mathscr{A})$. Consequently, it is easy to use Theorem 1 rather than Theorem 2 whenever the condition $A_1 \subseteq \cdots \subseteq A_k$ obtains as the following section illustrates.

3. An application of the main result

We specialize the notation of the previous sections by taking $\Theta = (0, \infty)$; $\mathscr{A}$ = Borel σ-field $\mathscr{B}$ in $(0, \infty)$ and that $X_1, \ldots, X_n$ is a random sample from the Dirichlet process P on $((0, \infty), \mathscr{B})$. The random variable X_i is randomly censored on the right by $Y_i \sim H_i$ for $i = 1, \ldots, n$. It is assumed that

$Y_1, \ldots, Y_n$ are independent, and

$(Y_1, \ldots, Y_n)$ is independent of $(P, X_1, \ldots, X_n)$. (A1)

Consequently, we observe only $(\boldsymbol{\delta}, \boldsymbol{Z}) = ((\delta_1, Z_1), \ldots, (\delta_n, Z_n))$ where

$$\delta_i = I_{[X_i \leq Y_i]} \quad \text{and} \quad Z_i = \min\{X_i, Y_i\}, \quad i = 1, \ldots, n. \tag{11}$$

It is desired to make decisions concerning F ($F(u) = P(0, u]$) when the data $X_1, \ldots, X_n$ is corrupted as described above. (For example, one might be interested in estimating F when

$$L(F, \hat{F}) = \int_0^\infty (F(u) - \hat{F}(u))^2 \mathrm{d}w(u) \quad \text{or} \quad L(F, \hat{F}) = \int_0^\infty |F(u) - \hat{F}(u)| \mathrm{d}w(u)$$

for some weight function w on $(0, \infty)$ or test for $F \leq F_0$ against the alternative $F \not\leq F_0$ under 0–1 loss function where F_0 is a known distribution function.) All these examples require the c.d. of F given $(\boldsymbol{\delta}, \boldsymbol{Z})$. We use Theorem 1 to obtain this c.d. which is stated below as Theorem 3.

Digressing slightly, we point out here that the model described above is a Bayesian analogue of the model extensively considered in connection with the analysis of survival data in medical studies. For example, see the papers by Gehan [3], and Kaplan and Meier [5], and the book by Gross and Clark [4].

In order to find the c.d. of P given $(\boldsymbol{\delta}, \boldsymbol{Z})$, we can assume (and will assume so), without loss of generality, that $\delta_1 = \delta_2 = \cdots = \delta_{n-k} = 1$, $\delta_{n-k+1} = \cdots = \delta_n = 0$ and that $Z_{n-k+1} \geq Z_{n-k+2} \geq \cdots \geq Z_n$. By Theorem 4 of [6], the c.d. of P given $(1, Z_1), \ldots, (1, Z_{n-k})$ is a Dirichlet process Q with parameter $\beta(\cdot) = \alpha(\cdot) + \sum_{j=1}^{n-k} I_{\cdot}(Z_j)$. Therefore, the needed c.d. of P given $(\boldsymbol{\delta}, \boldsymbol{Z})$ is equivalent to obtaining the c.d. of Q given $(0, Z_{n-k+1}), \ldots, (0, Z_n)$.

From a result on conditional distributions (for example, see (2.9.18) of Wilks [8]), it can be easily seen that for any measurable partition $A_1, \ldots, A_l$ of $(0, \infty)$,

$$\begin{aligned}
&\mathbf{P}\{Q(A_1) \leq a_1, \ldots, Q(A_l) \leq a_l \mid (0, z_{n-k+1}), \ldots, (0, z_n)\} = \\
&= \lim_{\varepsilon \downarrow 0} (\mathbf{P}\{Q(A_1) \leq a_1, \ldots, Q(A_l) \leq a_l;\ \delta_j = 0,\ z_j - \varepsilon < Z_j \leq z_j \qquad (12)\\
&\qquad \text{for } j = n-k+1, \ldots, n\} / \mathbf{P}\{\delta_j = 0,\ z_j - \varepsilon < Z_j \leq z_j \text{ for } j = n-k+1, \ldots, n\})
\end{aligned}$$

provided the rhs exists for $0 \leq a_i \leq 1$ for $i = 1, \ldots, l$ where $\mathbf{P}$ stands for probability operation. It now easily follows from the above formula that

$$\text{lhs of (12)} = \mathbf{P}\{Q(A_1) \leq a_1, \ldots, Q(A_l) \leq a_l \mid X_j \geq z_j \text{ for } j = n-k+1, \ldots, n\} \tag{13}$$

almost everywhere provided

$$H_j \text{ is absolutely continuous or discrete for } j = n-k+1, \ldots, n. \tag{A2}$$

Thus (13) shows that the c.d. of P given $(\boldsymbol{\delta}, \boldsymbol{Z})$ is equal to the c.d. of Q given $X_{n-k+j} \in A_j = [Z_{n-k+j}, \infty)$ for $j = 1, \ldots, k$. By assumption on Z_j,

$$A_1 \subseteq A_2 \subseteq \cdots \subseteq A_k. \tag{14}$$

Hence with the notation given above, (14) and Theorem 1 provide the following theorem.

Theorem 3. *Let H_j be absolutely continuous or discrete for $j = 1, \ldots, n$ and let* (A1) *hold. Then the* c.d. *of P given $(\boldsymbol{\delta}, \boldsymbol{Z})$ is a mixture of Dirichlet processes with transition measure $\beta(\cdot) + \sum_{j=1}^{k-1} \mu_j(\cdot) + I_\cdot(u)$ and with mixing measure μ_k where $\beta, \mu_1, \ldots,$ and μ_k are given by*

$$\beta(B) = \alpha(B) + \sum_{j=1}^{n-k} I_B(Z_j), \tag{15}$$

$$\beta([Z_{n-k+1}, \infty))\mu_1(B) = \beta(B \cap [Z_{n-k+1}, \infty)), \tag{16}$$

and

$$\mu_l(B) = \frac{\beta(B \cap [Z_{n-k+l}, Z_{n-k+l-1}))}{\beta([Z_{n-k+l}, \infty)) + l - 1} + \sum_{j=1}^{l-1} \frac{\beta(B \cap [Z_{n-k+j}, Z_{n-k+j-1}))}{\beta([Z_{n-k+j}, \infty)) + j - 1} \prod_{i=j}^{l-1} \left\{ \frac{\beta([Z_{n-k+i}, \infty)) + i}{\beta([Z_{n-k+i+1}, \infty)) + i} \right\} \tag{17}$$

for $l = 2, \ldots, k$.

The model described above can also be used for the Bayesian analysis of random transition times of Baboons (see Wagner and Altmann [7]).

References

[1] C.E. Antoniak, Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems, Ann. Statist. 2 (1974) 1152–1174.

[2] T.S. Ferguson, A Bayesian analysis of some nonparametric problems, Ann. Statist. 1 (1973) 209–230.

[3] E. Gehan, Estimating survival functions from the life table, J. Chron. Dis. 21 (1969) 629–644.

[4] A.J. Gross and V.A. Clark, Survival distributions. Reliability applications in Biomedical Sciences (Wiley, New York, 1975).

[5] E. Kaplan and P. Meier, Nonparametric estimation from incomplete observations, J. Am. Stat. Assoc. 53 (1958) 457–481.

[6] V. Susarla and J. van Ryzin, Nonparametric Bayesian estimation of survival curves from incomplete observations, J. Am. Stat. Assoc. 61 (1976) 897–902.

[7] S.S. Wagner and S.A. Altmann, What times do the baboons come down from the trees? (An estimation problem), Biometrics 29 (1973) 623–635.

[8] S.S. Wilks, Mathematical Statistics (Wiley, New York, 1962).